

Self-Reported Functional Status as Predictor of Observed Functional Capacity in Subjects with Early Osteoarthritis of the Hip and Knee: A Diagnostic Study in the CHECK Cohort

H. J. Bieleman · M. F. Reneman · M. W. van Ittersum ·
 C. P. van der Schans · J. W. Groothoff · F. G. J. Oosterveld

Published online: 26 June 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract *Objectives* Patients with hip or knee osteoarthritis (OA) may experience functional limitations in work settings. In the Cohort Hip and Cohort Knee study (CHECK) physical function was both self-reported and measured performance-based, using Functional Capacity Evaluation (FCE). Relations between self-reported scores on SF-36 and WOMAC (Western Ontario and McMaster Arthritis Index, function scales) and FCE performance were studied, and their diagnostic value for clinicians in predicting observed physical work limitations was assessed. *Methods* Ninety-two subjects scored physical function on SF-36 (scale 0–100, 100 indicating the best health level)

and WOMAC (scale 0–68, 68 indicates maximum restriction) and performed the FCE. Correlations were calculated between all scores. Cross-tables were constructed using both questionnaires as diagnostic tests to identify work limitations. Subjects lifting <22.5 kg on the FCE-test ‘lifting-low’ were labeled as having physical work limitations. Diagnostic aspects at different cut-off scores for both questionnaires were analysed. *Results* Statistically significant correlations (Spearman’s ρ 0.34–0.49) were found between questionnaire scores and lifting and carrying tests. Results of a diagnostic cross-table with cut-off point <60 on SF-36 ‘physical functioning’ were: sensitivity 0.34, specificity 0.97 and positive predictive value (PV+) 0.95. Cut-off point ≥ 21 on WOMAC ‘function’ resulted in sensitivity 0.51, specificity 0.88 and PV+ 0.88. *Conclusion* Low self-reported function scores on SF-36 and WOMAC diagnosed subjects with limitations on the FCE. However, high scores did not guarantee performance without physical work limitations. These results are specific to the tested persons with early OA, in populations with a different prevalence of limitations, different diagnostic values will be found. FCE may be indicated to help clinicians to assess actual work capacity.

H. J. Bieleman (✉) · F. G. J. Oosterveld
 Expertise Center Health, Social Care and Technology,
 Saxion University of Applied Sciences, P.O. Box 70.000,
 7500 KB Enschede, The Netherlands
 e-mail: h.j.bieleman@saxion.nl

M. F. Reneman
 Center for Rehabilitation, University Medical Centre Groningen,
 University of Groningen, Groningen, The Netherlands

M. W. van Ittersum · C. P. van der Schans
 Research and Innovation Group in Health Care and Nursing,
 Hanze University of Applied Sciences, Groningen,
 The Netherlands

H. J. Bieleman · M. F. Reneman · M. W. van Ittersum ·
 J. W. Groothoff
 Graduate School for Health Research, University Medical Center
 Groningen, University of Groningen, Groningen,
 The Netherlands

H. J. Bieleman · M. F. Reneman · M. W. van Ittersum ·
 J. W. Groothoff
 Department of Health Sciences, University Medical Center
 Groningen, University of Groningen, Groningen,
 The Netherlands

Keywords Osteoarthritis · Knee · Hip · FCE ·
 Physical function · Self-report

Introduction

Osteoarthritis (OA) of the hips and the knees is considered a major disabling disorder due to its restricting effect on mobility. While most prevalent in the elderly, recent publications demonstrated that younger people of working age may also be affected [1–3]. Disability at work depends on

the functional capacity of the person and on the physical, mental and social demands of the job. There is little information on physical function in relation to physical job demands for people with OA. Most studies focus on activities of daily life (ADL) limitations in the more advanced stages of the disorder in elderly people. Functional status in hip and knee OA generally deteriorates slowly [4]. It is feasible that in the early stages a high physical load during work may result in pain and functional limitations of workers. These people may have little or no limitations in ADL that are less demanding than their work. Reports on work limitations in degenerative joint disease are scarce [5].

Limitations in ADL are often measured with validated self-report instruments such as the 36-item Short-form Health Status Survey (SF-36 [6], generic) or Western Ontario and McMaster University Arthritis Index (WOMAC [7], arthritis specific). These instruments focus on perceived limitations, whereas performance based tests of functional capacity focus on observed test behavior. Functional Capacity Evaluations (FCE) are applied in specific contexts as pre-job screening, work rehabilitation and assessment of disability claims [8, 9]. The tests are physically demanding and take several hours to complete the full protocol. The validity of self-report and performance-based instruments is still under debate [10–12]. Terwee et al. [13] concluded that information on measurement properties of many performance-based methods for people with OA is incomplete, which makes it difficult to select an appropriate method. The psychometric properties of FCE have been described for healthy subjects and subjects with low back pain [14–16]. Reneman et al. [17] studied the concurrent validity of an FCE and self-reports on disability in relation to chronic low back pain. They found poor to moderate correlations between FCE results and outcomes of the low back related self-reported disability.

The Cohort Hip and Cohort Knee (CHECK) study [18] aims to study the course of early OA of the hip and the knee in people between 45 and 65 years (at inclusion). The course of impairments, disabilities and problems with social participation due to hip and knee complaints will be described. To cover a spectrum of biopsychosocial variables, a set of generic methods and instruments is used. We examined the potential use of two of these methods (self-report questionnaires) for predicting functional limitations on an FCE-battery. FCEs have been criticized because of the burden of testing, both for patients and clinicians. A good solution would be to develop a clinical rule to indicate if and when an FCE is needed to assess functional capacity for work. This rule would be helpful for general practitioners, rheumatologists, occupational physicians and physical therapists. Therefore, the objectives of this study were:

1. To describe the relation between on one hand the scores on SF-36 ‘physical function’ and WOMAC ‘function’ and on the other hand performance on a Functional Capacity Evaluation.
2. To determine the optimal cut-off point for the use of self-reports as diagnostic test to identify work limitations.
3. To study the diagnostic properties and diagnostic values of SF-36 and WOMAC in predicting limited functional capacity on the FCE.

Methods

Design

This study was a cross-sectional study in a sample of subjects participating in the CHECK cohort, a multi-centre longitudinal study on early OA ($n = 1002$) [18]. After inclusion in the cohort all subjects received a comprehensive questionnaire, composed from several validated questionnaires. All subjects from the CHECK-centres Groningen and Enschede ($n = 153$) were additionally invited to participate in this study in which the ability to perform work related activities was assessed with a Functional Capacity Evaluation.

Subjects

Inclusion criteria for the CHECK cohort were hip and/or knee complaints for which the subject visited the general practitioner no longer than 6 month ago and that were not attributed to direct trauma or other disorders. The age of the subjects was between 45 and 65 years. Exclusion criteria were the presence of inflammatory rheumatic disorders, joint prosthesis (hip and knee), previous joint trauma and serious co morbidity. All participants provided written informed consent before entering the study, and the Medical Ethical Board of hospital ‘Medisch Spectrum Twente’ in Enschede, The Netherlands, approved the study.

Measurements

Performance based outcome measures: The WorkWell Systems Functional Capacity Evaluation (WWS FCE) [19] was used to assess subjects’ work capacity. 22 tests, including all those that cause load bearing to the hips and the knees, were selected from the standardized 2-day WWS FCE protocol. These tests aim to record maximal capacity with regards to strength, endurance or speed. Providing the test leader judged the tests to be performed safely, subjects were asked to continue to a higher load level (five repetitions per level). The static endurance tests were continued

until a preset limit was reached. The subject was free to end any test at any moment, for example, because of discomfort or pain. Preceding the FCE tests subjects' age and sex were registered and the following measurements were performed: length, weight, Body Mass Index (BMI), location of the complaint (hip/knee/both and left/right/both).

Self-report outcome measures: The SF-36 and the WOMAC (Dutch versions) were used. The SF-36 [6] is a validated 36-item questionnaire that measures eight domains of health; in this study the scale for 'physical functioning' was used (containing 10 items with a 3 point Likert Scale, leading to a transformed score range of 0–100 in 20 steps of 5 points, 100 indicating the highest level of functioning). The WOMAC [7] is a validated self-administered questionnaire for patients with hip or knee OA, consisting of 24 questions categorized in subscales of pain, stiffness and function. In this study the 'function' scale was included in the analyses (17 items, 5 point Likert Scale, score range 0–68 in 68 steps, 68 indicating maximal restrictions in function).

Diagnostic cross-table: Analogous to diagnostic tests for diseases 2×2 cross-tables were constructed for disease presence (yes/no) and diagnostic test result (positive/negative). In our cross-tables the presence of observed work limitations in the FCE was related to scores on the self-report questionnaires. To split the subjects in a group with work limitations and a group without work limitations, criteria from the Dictionary of Occupational Titles (DOT [20]) were used. The DOT categorises physical job demands into five categories, which are mainly based on the amount of weight to be lifted in the job. Subjects only able to perform work tasks which lay within the lowest physical levels of activity, classified as sedentary or light tasks (lifting occasionally up to 22.5 kg, based on the FCE test 'lifting low') were labeled as having 'work limitations'. Those who were able to perform medium, heavy or very heavy work (lifting occasionally 22.5 kg and more) were considered to have 'no work limitations'. Questionnaire results reflecting self-reported restrictions in physical function (scores below a chosen cut-off value for SF-36 and scores over a WOMAC cut-off point) indicated a positive test result, the remaining scores indicated a negative result. In summary, a cross-table was constructed to evaluate the potential diagnostic value of the physical function subscales of SF-36 and WOMAC (self-reports) in predicting functional work limitations on the FCE (performance test).

Protocol

Questionnaires were filled in on inclusion into the cohort. FCE was performed after subjects gave informed consent to participate in this spin-off study (additional to the

cohort). As a result there was a time lapse between the self-reporting and the FCE. Tests were led by 4th year Physical Therapy students who received a 1-day training in the procedure and the execution of the tests. They were supervised by the research team. Testers were blinded for the self-report outcomes and the criteria for interpretation (22.5 kg).

Statistical Analysis

Descriptive statistics were performed on the results from FCE, SF-36 'physical function' and WOMAC 'function'. Correlations between FCE performance and questionnaire scores were assessed using Spearman rank correlation coefficients. Bonferroni procedures [21] were applied to reduce type I error, adjustment for 44 comparisons at $\alpha = 0.05$ resulted in the use of $P < 0.001$ as level of significance.

Frequency tables of 'lifting low' performance for different SF-36 scores and WOMAC scores were used to construct cross tables for a series of cut-off points. Diagnostic properties and diagnostic values of the tests (see the text box for an introduction) were calculated for each cut-off point.

A brief introduction to diagnostic properties and values

Sensitivity (Se) is the probability of a positive test outcome given that the disorder (in this study: work limitations) is present, specificity (Sp) is the probability of a negative test outcome given that work limitations are not present. Of practical importance for clinicians are the positive predictive value (PV+), this is the probability that an individual has work limitations in case of a positive test outcome, and the negative predictive value (PV−), this is the probability that an individual does not have work limitations in case of a negative test outcome. However, both PV+ and PV− are affected by the prevalence of work limitations in the studied population.

Statistical as well as clinical criteria were used to determine the optimal cut-off point for SF-36 and WOMAC scores that indicated a positive test. Results for the chosen cut-off points were displayed in scatter plots with scores on questionnaire versus FCE performance on 'lifting low'. To match the plots with the quadrants of the diagnostic cross tables the SF-36 scores on the y-axis were inverted: 0 was put on top of the y-axis, because low scores indicate a positive diagnostic test outcome.

Since only the 'lifting low' test was used to determine the cut-off points of the self-reports, we subsequently examined whether applying these cut-off scores to the other FCE tests would also clearly divide the subjects in low and high performers. This was done by testing the differences in performances on all the other FCE tests

Table 1 Subject characteristics of FCE-participants, non-participants and the rest of the cohort

Variable	FCE-participants (<i>n</i> = 92)		Non FCE-participants (<i>n</i> = 61)		Others in cohort (<i>n</i> = 849)	
	Mean (SD)	Median (range)	Mean (SD)	Median (range)	Mean (SD)	Median (range)
Age (years)	56 (4.9)	56 (47–65)	55 (5.8)	55 (45–65)	56 (5.2)	56 (45–65)
Female sex (%)	84		87		78	
Work participation (%)	48		50		46	
BMI (kg/m ²)	26.0 (4.5)	25.1 (13–40)	26.2 (3.7)	25.3 (21–40)	26.2 (4.1)	25.5 (15–49)
SF-36 physical functioning ^a	71.2 (21.6)	75 (5–100)	74.3 (16.5)	75 (25–100)	75.1 (16.7)	80.0 (5–100)
WOMAC function ^b	18.1 (12.1)	15.0 (0–49)	16.5 (11.7)	16.0 (0–53)	15.7 (11.6)	13.0 (0–56)

^a On a scale of 0 (worst situation) to 100 (best situation)

^b On a scale of 0 (best situation) to 68 (worst situation)

between persons with a positive test and those with a negative test. Independent samples *t*-tests were used on the manual material handling tests; Mann–Whitney tests were used on the other tests, because of ceiling and criterion effects. The level of significance (α) was chosen at 0.05.

Results

Subjects

Ninety-two of the 153 invited CHECK participants were enrolled in this study (79 women, 13 men). Of this sample, 59 had complaints of the hip(s) as well as the knee(s). Subjects' characteristics are described in Table 1. They were very similar to the other 849 subjects in the cohort and to the 61 non-participants with regards to age, sex, body mass index, work participation and scores on physical function scales of SF-36 and WOMAC.

Study Objective 1: Correlations

Spearman's rho (ρ) for correlations between the scores on SF-36, WOMAC and FCE are presented in Table 2. WOMAC correlations were negative where SF-36 correlations were positive because at the WOMAC higher scores indicate more restrictions. The highest correlation was found between the two self-report instruments. Correlations between self-reports and nearly all manual material handling FCE tests were statistically significant with ρ -values ranging from 0.34 to 0.49. Correlations with most of the other FCE-tests were not statistically significant. Results for the stair climbing test (10 × 10 stairs) were not presented because 34 subjects reached the preset heart rate safety limit (85% of maximal heart rate) and had to end the test prematurely.

Table 2 Spearman rank correlation coefficients for SF-36, WOMAC and FCE tests

	SF-36 (physical function)	WOMAC (function)
WOMAC function	.70*	
Manual material handling (kg)		
Lifting low	.37*	–.37*
Lifting high	.32	–.39*
Carry short	.36*	–.39*
Carry long, two hands	.34*	–.43*
Carry long, R hand	.46*	–.47*
Carry long, L hand	.38*	–.49*
Push static	.20	–.34
Pull static	.31	–.37*
Static posture, endurance (s)		
Static overhead work	.13	–.32
Static bent work	.26	–.29
Kneeling	.33	–.45*
Squat	.23	–.18
Dynamic movements, speed (s)		
Crawling	–.21	.24
Dynamic bent work	–.20	.30
Repetitive squats	–.27	.36*
Stand L repetitive rotation	–.13	.19
Stand R repetitive rotation	–.12	.18
Sit L repetitive rotation	–.04	.12
Sit R repetitive rotation	–.11	.18
Ladder	–.33	.30
Dynamic movements, endurance (m)		
Shuttle walk	.25	–.39*

* $P < 0.001$

Study Objective 2 and 3: Cut-off Points and Diagnostic Values

In Table 3 the diagnostic qualities at different cut-off points are presented of both SF-36 'physical function' and

Table 3 Properties of SF-36 ‘physical function’ and WOMAC ‘function’ as a diagnostic test for work limitations, at different cut-off points

Self-report instrument	Cut-off point	Positive and negative tests (\pm)	Sensitivity	Specificity	Likelihood ratio of positive test
SF-36 physical function	55	17/75	0.27	0.97	8.9
	60	21/71	0.34	0.97	11.1
	65	26/66	0.41	0.94	6.7
	70	33/59	0.47	0.85	3.1
	75	42/50	0.59	0.79	2.8
	80	47/45	0.64	0.73	2.4
	85	55/37	0.73	0.64	2.0
	90	68/24	0.81	0.39	1.3
WOMAC function	25	28/64	0.41	0.88	3.4
	24	30/62	0.44	0.88	3.6
	23	31/61	0.46	0.88	3.8
	22	32/60	0.47	0.88	3.9
	21	34/58	0.51	0.88	4.2
	20	38/54	0.54	0.82	3.0
	19	39/53	0.56	0.82	3.1
	18	40/52	0.56	0.79	2.6
	17	43/49	0.59	0.76	2.4
	16	45/47	0.61	0.73	2.2

WOMAC ‘function’, in relation to work limitations (the defined ‘disease’).

The table illustrates that, as in every diagnostic test, shifting the cut-off point resulted in a trade-off between sensitivity and specificity. For SF-36 a cut-off point of <60 points was chosen, because at this score the highest specificity (0.97) is reached in combination with a high likelihood ratio for a positive test (11.1); 21 subjects (23%) were tested ‘positive’. For WOMAC a cut-off point of ≥ 21 was chosen, which gave lower specificity and higher sensitivity compared to SF-36. This cut-off point resulted in 34 subjects (37%) with a positive test.

In Fig. 1 scatter plots of the results of all subjects are presented in combination with cross-tables with the diagnostic values at the chosen cut-off points. The self-report scores predicted low performance on the FCE-test ‘lifting low’ for 20 out of 21 positive tests on the SF-36 (positive predictive value, $PV+ = 0.95$) and for 30 out of 34 positive tests on the WOMAC ($PV+ = 0.88$). The $PV-$ for SF-36 and WOMAC were 0.45 and 0.50, respectively.

In Table 4 the performances on all the FCE tests are compared for subjects with positive and negative diagnostic tests. These results indicate that on manual material handling tests persons with negative tests (high self-reported function) handled heavier weights. All differences in test results were statistically significant.

For static posture tests the results were mixed. Although not all of them were statistically significant the tendency was for both SF-36 and WOMAC that subjects with

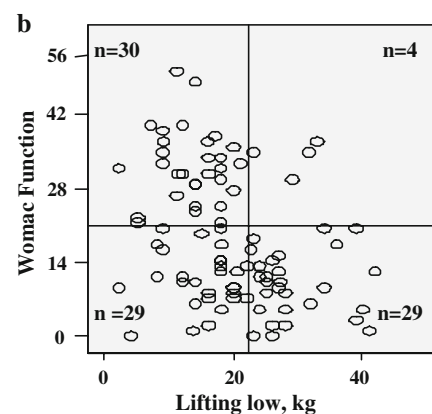
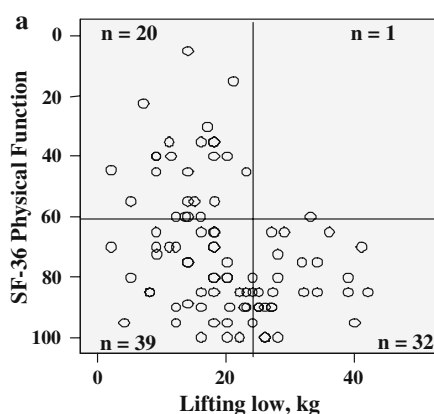
negative tests demonstrated higher endurance. Most of the dynamic tests did not show significantly different results, although the group with negative tests performed faster on average. On the shuttle walk test persons with negative diagnostic tests walked longer distances. In summary the group with good self-reported function performed better on all FCE tests.

Discussion

The main objectives of our study on persons with early OA of the hip and the knee were to describe relations between scores on the function scales of SF-36 and WOMAC and performance on the FCE and to determine the diagnostic value of these scales in predicting limited capacity on the FCE. If these questionnaires demonstrate predictive value in identifying physical work limitations they can help clinicians to decide whether or not an FCE is indicated to evaluate physical work capacity.

The invitation to voluntarily participate in this study could have introduced selection bias, if for example people with a higher physical capacity were more willing to perform the demanding tests. Our results however indicated that the subjects were similar to the non-participants on the compared variables. Neither were there any differences in comparison to the rest of the cohort with respect to age, sex, work participation, body mass index and SF-36 and WOMAC scores. These scores indicated that most of our

Fig. 1 a Scatter plot for lifting performance versus SF-36 'physical function' with cut-off scores indicated; to match the plots with the quadrants of the diagnostic cross tables, the SF-36 scores on the y-axis were inverted (0 on top of the y-axis); corresponding cross table + diagnostic values. **b** Scatter plot for lifting performance versus WOMAC 'function' with cut-off scores indicated; corresponding cross table + diagnostic values



SF-36:		Work Limitations		
Physical Function, cut-off point <60	<60: + ≥60: -	<22.5 kg + -		
		≥22.5 kg + -		
		20	1	21
		39	32	71
		59	33	92

Prevalence = 59/92 = 0.64

Sensitivity = 20/59 = 0.33 (95%CI: 0.23-0.43)

Specificity = 32/33 = 0.97 (95%CI: 0.94-1.0)

Predictive Value + = 20/21 = 0.95 (95%CI: 0.91-0.99)

Predictive Value - = 32/71 = 0.45 (95%CI: 0.35-0.55)

Likelihood Ratio + = 11.2 (95%CI: 5.9-21.3)

Likelihood Ratio - = 0.68 (95%CI: 0.49-0.94)

WOMAC:		Work Limitations			
Function, cut-off point ≥21	≥21: + <21: -	<22.5 kg +		≥22.5 kg -	
		30	4		34
		29	29		58
		59	33		92
Prevalence = 59/92 = 0.64					
Sensitivity = 30/59 = 0.51 (95%CI: 0.41-0.61)					
Specificity = 29/33 = 0.88 (95%CI: 0.81-0.95)					
Predictive Value + = 30/34 = 0.88 (95%CI: 0.81-0.95)					
Predictive Value - = 29/58 = 0.50 (95%CI: 0.40-0.60)					
Likelihood Ratio + = 4.2 (95%CI: 2.32-7.60)					
Likelihood Ratio - = 0.6 (95%CI: 0.33-1.09)					

subjects, included as having early OA, were in relatively good self-reported health.

The correlations between the scores on questionnaires and the performance on the FCE varied in a logical manner that provides construct validity to subtests of the FCE. A number of questionnaire items correspond almost literally with FCE items (for example lifting or carrying groceries, kneeling/stooping, walking). Other items refer to activities that are not in the FCE protocol (for example bathing or dressing), while some FCE tests (for example repetitive movements) do not match with questionnaire items. Furthermore, the relation between self-reported functional status and observed performance must have been influenced by other than physical factors. Both physical and psychological factors have been identified as having influence on the functional status with regard to mobility of older people with OA [22–25]. FCE tests that require strength showed the highest correlations with the self-reports. An explanation may be that these tests put the

highest mechanical loads on the hips and knees, resulting from the combination of body movements and the weights lifted or carried. Self-reported disability because of pain or discomfort was expressed clearly on these tests. In the other tests speed or endurance were more called on than strength and factors such as dexterity or willingness to continue may have become decisive.

Similar to diagnostic tests for diseases we constructed diagnostic cross-tables. The aim of this action was to explore whether those subjects who showed work limitations on the (physically demanding) FCE could be identified based on their (easily obtained) self-reported functional score. Although we performed a cross-sectional study we used the term 'prediction' to indicate whether questionnaire scores gained useful information about subsequently observed performance. Our choice of the FCE test 'lifting low' as criterion for work limitations was based on the DOT-system in which lifting of weights is regarded as a critical job demand. The figure of 22.5 kg corresponds

Table 4 Comparison of mean or median results on the FCE tests for groups SF+ and SF– and for WOMAC+ and WOMAC–, tested with independent *t*-tests (manual material handling) or Mann–Whitney tests (others)

	Mean SF+, <i>n</i> = 21	Mean SF–, <i>n</i> = 71	Mean diff.	<i>P</i>	Mean WOMAC+, <i>n</i> = 34	Mean WOMAC–, <i>n</i> = 58	Mean diff.	<i>P</i>
Manual material handling (kg)								
Lifting low	13.8	21.5	7.7	.001	15.2	22.4	7.2	.000
Lifting high	5.8	10.7	4.8	.000	6.8	11.2	4.4	.000
Carry short	13.2	21.9	8.7	.000	14.6	23.0	8.4	.000
Carry 2 hand	16.2	24.4	8.1	.002	17.2	25.6	8.4	.000
Carry right	16.4	23.4	7.0	.001	16.4	25.0	8.6	.000
Carry left	15.4	23.1	7.7	.001	15.8	24.6	8.8	.000
Push static	21.0	28.5	7.5	.010	21.5	30.1	8.5	.000
Pull static	24.0	35.9	11.9	.003	26.4	37.4	11.0	.000
	Median SF+	Median SF–		<i>P</i>	Median WOMAC+	Median WOMAC–		<i>P</i>
Static posture, endurance (s)								
Static overhead	145	166		.353	144	174		.006
Static bent	191	339		.006	231	378		.005
Kneeling	146	300		.001	236	300		.001
Squat	60	60		.017	60	60		.099
Dynamic movements, speed (s)								
Crawling	51	48		.083	54	43		.011
Bent dynamic	60	54		.114	61	53		.018
Repeated squat	53	49		.102	53	48		.007
Stand L repetitive rotation	89	81		.024	86	83		.068
Stand R repetitive rotation	86	79		.105	84	79		.105
Sit L repetitive rotation	90	84		.247	88	84		.350
Sit R repetitive rotation	91	85		.249	91	85		.064
Ladder	143	113		.013	136	111		.009
Dynamic movements, endurance (m)								
Shuttle walk	250	330		.011	250	330		.000

with the limit between light and medium physical demands (DOT) and also equals the recommended weight limit of the NIOSH guideline [26] that claims to be safe for 99% of men and 75% of women in an ideal lifting situation. We considered the DOT and the NIOSH guidelines as widely accepted and best available evidence for choosing a criterion. Applying this 22.5 kg limit the prevalence of work limitations in our subjects was 64%. Since 85% of our subjects were women with a mean age of 56 and less than 50% of them were in paid work, this result seems plausible.

In our cross-table we have chosen a cut-off point of <60 points on the SF-36 subscale physical function as criterion for a positive diagnostic test. This choice was based on a combination of parameters, i.e. the likelihood ratio for a positive test (LR+), the high predictive value of a positive test, the high specificity, and a useful number of positive tests.

The diagnostic cross-table enabled us to predict low performance on the FCE-test ‘lifting low’ based on poor self-reported physical function for 21 of our 92 subjects, with 95% ‘true positive’ outcomes. The LR+ of 11.2 indicated that this positive test outcome increased the odds of subjects demonstrating work limitations on the FCE from the base rate of 59/33 to 20/1. The osteoarthritis specific WOMAC was cut-off at a score of ≥ 21 points (on the 0–68 ‘function’ scale). The use of this cut-off point identified 34 subjects with a positive test (poor self-reported function) and resulted in 88% ‘true positive’ outcomes. Compared to SF-36 the WOMAC identified 13 more subjects with work limitations at the cost of a 7% decrease in certainty of this positive diagnosis. Apparently the strength of both questionnaires lies in its positive predictive value to identify subjects with work limitations in the early stage of the OA.

The use of the FCE-test ‘lifting low’ as criterion for work limitations was supported by the outcomes of applying the same diagnostic criterion (a SF-36 ‘physical functioning’ score <60 or WOMAC ‘function’ ≥ 21) to the other ‘manual material handling’ tests of the FCE. Although we did not present them, the resulting scatter plots and cross tables were very similar. We concluded that these scores indeed predict physical work limitations, especially where lifting and carrying were critical job demands. These are the same FCE tests that showed significant correlations with self-report scores (Table 2).

The negative predictive value of the questionnaire scores in our diagnostic cross-table was low, due to the many subjects with good self-reported functional status who nevertheless demonstrated low FCE-scores. The questionnaires capture limitations in a range of ADL but do not refer sufficiently to specific work related activities. The strength of SF-36 and WOMAC lies therefore not in selecting people that are capable to perform heavier work; for that aim additionally the FCE can be used. In populations with a different prevalence of work limitations, the PV+ and PV– will be different; for example in a population of healthy workers with a lower prevalence of work limitations, a lower PV+ and a higher PV– are expected.

A limitation of this study was that due to the inclusion procedure an average time lapse of 5 months arose between answering of the questionnaires and participation in the FCE. We assumed both measurements to be relatively stable at the start of our cohort. Van Dijk et al. [4] concluded in her review that functional status in hip and knee OA deteriorates slowly in the first 3 years. FCE measurements do show a high test-retest reliability but also some natural variation [15, 16] within the individual. The FCE data of the first follow-up measurement (T_1 , 1 year later) however do not indicate performance changes compared to the baseline measurement.

Our diagnostic cross-tables demonstrated that scores that indicate worse self-reported functional status were related to low performance on a Functional Capacity Evaluation in early osteoarthritis of the hip and the knee. We agree with Vignon et al. [27] that in general health care practice awareness must be stimulated for the relation between hip and knee complaints of younger people and their work capacity. Patients with physically demanding work should be advised to visit the occupational physician and/or the Human Resources Management staff of their employer to discuss the opportunities for work adaptations. In the setting of occupational health care the use of an FCE in addition to self-reports is advised for a more specific assessment of work capacity. Also more occupation specific questionnaires or surveys should be selected or developed and translated in different languages. These should also cover mental and social work aspects. Follow-

up studies on work limitations in OA will be done in the CHECK cohort.

In conclusion, in subjects with early OA low self-reported physical function scores on SF-36 and WOMAC both demonstrated good diagnostic value as tests for limitations on the FCE. However, the diagnostic values are disorder specific and therefore in populations with a different prevalence of limitations, different diagnostic values will be found. Depending on the level of accuracy needed, self-reports may be sufficient to assess limitations in physical function. High self-reported scores did not guarantee performance without physical work limitations. Therefore, an FCE may be indicated to help clinicians to assess actual work capacity.

Acknowledgments “CHECK is funded by the Dutch Arthritis Association on the lead of a steering committee comprising 16 members with expertise in different fields of OA chaired by Prof. J. W. J. Bijlsma and coordinated by J. Wesseling, MSc. Involved are: Academic Hospital Maastricht; Erasmus Medical Center Rotterdam; Jan van Breemen Institute/VU Medical Center Amsterdam; Kennemer Gasthuis Haarlem; Martini Hospital Groningen/Allied Health Care Center for Rheumatology and Rehabilitation Groningen; Medical Spectrum Twente Enschede/Twenteborg Hospital Almelo; St. Maartenskliniek Nijmegen; Leiden University Medical Center; University Medical Center Utrecht and Wilhelmina Hospital Assen.” Grants for this study were received from NutsOhra Fonds and Mobiliteitsfonds HBO. The authors are grateful to the subjects who actively participated in the study and to the students that led the FCE tests. Also they thank Ans Oostveen, rheumatologist, and Anita Mooij and Annet ter Avest, research nurses of the Medisch Spectrum Twente and the Ziekenhuisgroep Twente, Wim Hilberdink, physical therapist in Groningen, and Janet Wesseling, CHECK-coordinator, for their contributions to the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Flugsrud GB, Nordsletten L, Espehaug B, Havelin LI, Meyer HE. Risk factors for total hip replacement due to primary osteoarthritis. A cohort study in 50034 persons. *Arthritis Rheum.* 2002;46(3):675–82.
2. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA. The work limitations questionnaire’s validity and reliability among patients with osteoarthritis. *J Clin Epidemiol.* 2002;55:197–208.
3. Fautrel B, Hilliquin P, Rozenberg S, Allaert F, Coste P, et al. Impact of osteoarthritis: results of a nationwide survey of 10,000 patients consulting for OA. *Joint Bone Spine.* 2005;72:235–40.
4. Van Dijk G, Dekker J, Veenhof C, van den Ende CHM. Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. *Arthritis Rheum (Ar C Res).* 2006;55(5):779–85.
5. Gobelet C, Luthi F, Al-Khodairy AT, Chamberlain MA. Work in inflammatory and degenerative joint diseases. *Disabil Rehabil.* 2007;29(17):1331–9.

6. McHorney CA, Ware JE, Racze AE. The MOS 36 item short-form health status survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247–63.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15:1833–40.
8. Harbin G, Olson J. Post-offer, pre-placement testing in industry. *Am J Ind Med*. 2005;47:296–307.
9. Wind H, Gouttebarga V, Kuijer PP, Frings-Dresen MHW. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living and sport: a systematic review. *J Occup Rehabil*. 2005;15:253–72.
10. Kivinen P, Sulkava R, Halonen P, Nissinen A. Self-reported and performance-based functional status and associated factors among elderly men: the Finnish cohort of the seven countries study. *J Clin Epidemiol*. 1998;51(12):1243–52.
11. Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. *Ann Rheum Dis*. 2006;65(Suppl III):iii16–21.
12. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip and knee. *Phys Ther*. 2006;86(11):1489–96.
13. Terwee CB, Mokkink LB, Steultjens MPM, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology*. 2006;45:890–902.
14. Reneman MF, Dijkstra PU, Westmaas M, Göeken LNH. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil*. 2002;12:269–75.
15. Lygren H, Dragesund T, Joensen J, Ask T, Moe-Nilssen RM. Test-retest reliability of the progressive isoinertial lifting evaluation (PILE). *Spine*. 2005;30(9):1070–4.
16. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, Göeken LNH. Test-retest reliability of the Isernhagen work systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehabil*. 2003;13(4):207–18.
17. Reneman MF, Jorritsma W, Schellekens JMH, Göeken LNH. Concurrent validity of questionnaire and performance based disability measurements in patients with chronic low back pain. *J Occup Rehabil*. 2002;12:119–29.
18. Wesseling J, Dekker J, Van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, et al. CHECK: cohort hip & cohort knee; similarities and differences with the OA initiative. *Ann Rheum Dis*. 2008. doi:10.1136/ard.2008.096164.
19. WorkWellSystems, FCE. Manual V2. Duluth; 2006.
20. Dictionary of Occupational Titles. 4th ed. US Department of Labor, Employment and Training Administration; 1991.
21. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
22. Maly MR, Costigan PA, Olney SJ. Determinants of self-report outcome measures in people with knee osteoarthritis. *Arch Phys Med Rehabil*. 2006;87:96–104.
23. Peters TJ, Sanders C, Dieppe P, Donovan J. Factors associated with change in pain and disability over time: a community-based prospective observational study of hip and knee osteoarthritis. *Br J Gen Pract*. 2005;55:205–11.
24. Jinks C, Jordan K, Croft P. Osteoarthritis as a public health problem: the impact of developing knee pain on physical function in adults living in the community: (KNEST 3). *Rheumatology*. 2007;46:877–81.
25. O'Reilly SC, Jones A, Muir KR, Doherty M. Quadriceps weakness in knee osteoarthritis: the effect on pain and disability. *Ann Rheum Dis*. 1998;57:588–94.
26. Waters TR, Putz-Anderson V, Garg A, Fine LJ. Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*. 1993;36:749–76.
27. Vignon E, Valat J, Rossignol M, Avouac B, Rozenberg S, et al. Osteoarthritis of the knee and hip and activity: a systematic international review and synthesis (OASIS). *Joint Bone Spine*. 2006;73:442–55.